

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 05-153581

(43)Date of publication of application : 18.06.1993

(51)Int.Cl.

H04N 7/14

G06F 15/62

G10L 3/00

H04N 7/13

(21)Application number : 03-318055

(71)Applicant : SEIKO EPSON CORP

(22)Date of filing : 02.12.1991

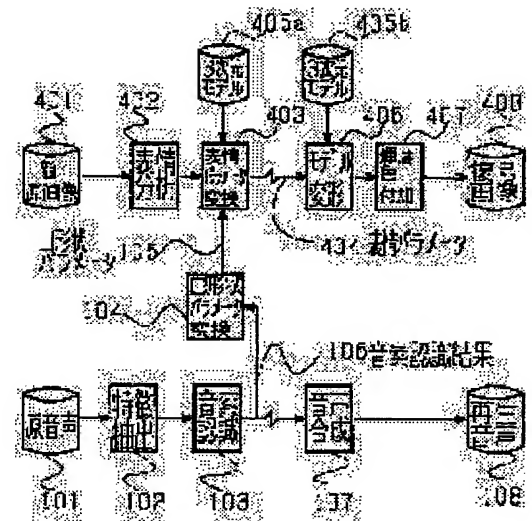
(72)Inventor : HORII KAZUYA

(54) FACE PICTURE CODING SYSTEM

(57)Abstract:

PURPOSE: To obtain the face picture coding system in which the processing is facilitated and the data compression efficiency is high.

CONSTITUTION: An face original picture 401 is given to an expression analysis section 402 and an expression parameter conversion section 403, from which an expression deformation parameter (except mouth) 404 is obtained with respect to a 3-dimension model 405a. On the other hand, an original voice signal 101 is given to a mouth shape parameter conversion section 104 via a characteristic extraction section 102 and a phoneme recognition section 103, in which the signal is converted into a mouth shape parameter 105, the 3-dimension model 405a is subject to model deformation 406 via the expression deformation parameter 404 and the mouth shape parameter 105, and a decoding picture 408 is obtained through the luminance and color addition processing 407.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平5-153581

(43)公開日 平成5年(1993)6月18日

(51)Int.Cl. ⁵	識別記号	片内整理番号	F I	技術表示箇所
H 0 4 N 7/14		8943-5C		
G 0 6 F 15/62	3 5 0	8125-5L		
G 1 0 L 3/00		Q 8946-5H		
H 0 4 N 7/13		Z 4228-5C		

審査請求 未請求 請求項の数2(全 7 頁)

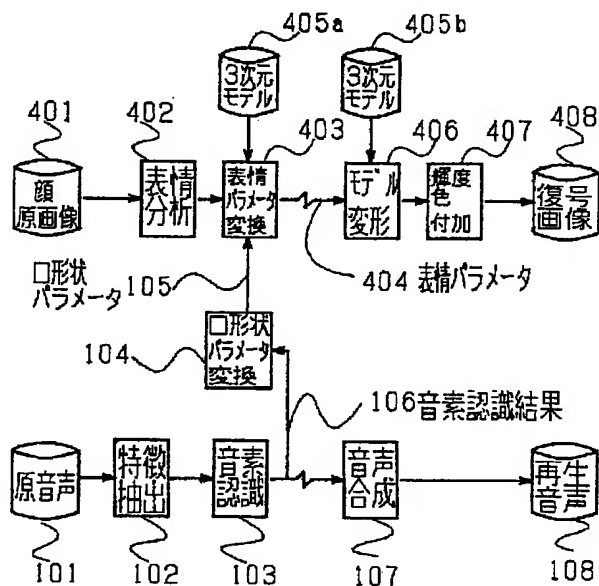
(21)出願番号	特願平3-318055	(71)出願人	000002369 セイコーエプソン株式会社 東京都新宿区西新宿2丁目4番1号
(22)出願日	平成3年(1991)12月2日	(72)発明者	堀井 和哉 長野県諏訪市大和3丁目3番5号セイコーエプソン株式会社内
		(74)代理人	弁理士 鈴木 喜三郎 (外1名)

(54)【発明の名称】 顔画像符号化方式

(57)【要約】

【目的】 本発明は、処理が容易で、データ圧縮効率の高い顔画像符号化方式を提供することを目的とする。

【構成】 顔原画像401は、表情分析402、表情パラメータ変換403を経て、3次元モデル405aに対す、口以外の表情変形パラメータ404が求められる。一方、原音声101は特徴抽出102、音素認識103を経て、口形状パラメータ変換104により、口形状パラメータ105に変換される。前記表情変形パラメータ404と口形状パラメータ105により、3次元モデル405bがモデル変形406され、輝度、色付加407の処理を経て、復号画像408が得られる。



【特許請求の範囲】

【請求項1】 符号化側、および、復号化側において、同様に、顔画像の2次元、あるいは、3次元モデルを所有し、符号化側では顔原画像から表情分析を行い、前記2次元、あるいは、3次元モデルの変形情報を出力し、復号化側では前記変形情報をもとに、2次元、あるいは、3次元モデルを変形させて画像を合成する顔画像符号化方式において、

顔原画像に付随する音声情報が存在する場合には、該音声情報より音素を認識して、3次元モデルにおける口、および、口周辺の変形情報として使用し、音声情報が存在しない場合には、原画像から認識した変形情報のみを使用して符号化を行うことを特徴とする顔画像符号化方式。

【請求項2】 前記音素認識結果から抽出した、顔画像における口、および、口周辺の変形情報を基に合成した画像と、入力された原画像との差分値から、2次元、あるいは、3次元モデルの変形情報を補正することを特徴とする請求項1記載の顔画像符号化方式。

【発明の詳細な説明】

【0001】

【産業上の利用分野】この発明は、画像通信、あるいは、画像蓄積等への応用を目的とした、顔画像データ圧縮のための画像符号化方式に関するものである。

【0002】

【従来の技術】近年、テレビ電話、テレビ会議等の画像の狭帯域通信、あるいは、蓄積メディア等への応用を目指して、画像データ圧縮のための画像符号化方式の開発、実用化が盛んに行われている。例えば、画像を小さいブロックに分割して、ブロック内画素を互いに無相関な軸へ変換することにより冗長性を取り除く変換符号化、あるいは、ブロック内の画素から構成されるベクトルを代表ベクトルに写像して、その代表ベクトルの番号を符号化データとするベクトル量子化等様々な符号化方式が存在する。しかしながら、前記、変換符号化、ベクトル量子化に代表される従来の符号化方式は、総括すると、画像における輝度、色、あるいは、色差などの空間的な変化そのものを忠実に再現することを目的とした、波形符号化に分類されるものであり、その符号化効率には限界がある。例えば、画像通信において、64Kbps (bps:ビットパーセカンド、1秒間に送ることのできるビット数)の伝送路で動画を送る場合、波形符号化では、一般のテレビ放送と比較して、動き、画質とも満足のいくものではない。そこで、最近、従来のデータ圧縮の考え方を大きく転換させ、高画質、高効率を目指した分析・合成符号化と呼ばれる画像符号化方式が研究されている。分析・合成符号化は波形符号化の限界を越えるものとして注目されはじめており、近年においては、特に、テレビ電話、テレビ会議等への応用を目指して、人間の顔画像の分析・合成符号化の研究が盛んに行

われている。顔画像の分析・合成符号化においては、符号化側と復号化側で、同じ顔画像の3次元モデルを所有し、符号化側では、各画素値の集合として与えられる顔画像データを分析して、3次元モデルに対する変形情報を抽出し、符号化データとする。一方、復号化側では、前記符号化データ、即ち、3次元モデルの変形情報に基づき、予め所有する3次元モデルを変形させ、輝度や色データを付加して、顔画像を再生する。

【0003】さて、従来、顔画像を対象としたこの種の分析・合成符号化方式の概念を示したものとしては、例えば、信学論 B-1 Vol. J72-B-1 No. 3 (1989) 相澤・原島・斉藤「構造モデルを用いた画像の分析合成符号化方式」に示されたものがある。図4は、前記引用例に示されている図を、本発明との比較を容易にするために修正して示したものである。図において、顔原画像401はカメラ、スキャナ等の画像入力装置から入力される人間の顔画像、表情分析402は、前記顔原画像401を入力データとし、表情を分析する。表情パラメータ変換403では、前記表情分析402の出力データに基づき、3次元モデル405aを基準とした時の表情パラメータ404を抽出し、符号化データとして出力する。一方、復号化側では、まず、受け取った符号化データ、即ち、表情パラメータ404に基づき、モデル変形406において3次元モデル405bを変形させる。さらに、輝度・色付加407では、変形後のモデル画像に対して、画素毎に輝度・色をつけ、表示するための復号画像408が生成される。

【0004】次に動作について説明する。

【0005】図4において、画像入力装置等から入力された人間の顔原画像401は、いわゆる、ビットマップ画像と呼ばれるものであり、画像を構成する各画素毎に画素値を持った、圧縮されていない状態のデジタル画像である。この顔原画像401は、次段の表情分析402において、まず、フィルタリング処理等を用いたノイズ除去、輝度や色や顔の大きさ等の正規化、顔輪郭線の抽出、あるいは、目、眉、鼻、口などの、顔を構成する各要素の抽出処理が行われる。次に、抽出された各要素の相対位置関係、輪郭線との相対位置関係、あるいは、各要素の大きさ等の情報を用いて、表情の分析が行われる。表情の種類としては、例えば、典型的なものとして、無表情(中立)、喜び、怒り、驚き、哀しみ、恐れ、嫌悪などがあるが、ここでは、「表情」を広義にとらえ、瞳の向きやまばたき、あるいは頭部全体の動き等についても含むものとする。ただし、現時点においては、前記の喜怒哀楽等の表情を認識することはむづかしい状況であるため、顔の各要素毎に基本となる動きパターンを数種類ずつ決め、その基本パターンを抽出しているのが現状である。たとえば、眉に関する基本パターンとしては、(1)眉の内側を上げる、(2)眉の外側を上げる、(3)眉を下げる、などがある。さて、上記の

ようにして求められた表情分析データは、次に表情パラメータ変換403において、3次元モデル405aの変形パラメータに変換される。3次元モデル405aは、ワイヤー（線分）によって形状を表現する、例えば、図5に示したようなワイヤーフレームモデルが使用される。各ワイヤーの始点、終点の座標は、予め、入力される顔原画像401に適合するように、拡大、縮小、平行移動、回転等によって変換されている。即ち、顔原画像401と図5のワイヤーフレームモデルは、重なるように変換されており、当然のことながら、復号化側に対しても同様の交換情報が送られて3次元モデル405bは、3次元モデル405aに全く等しく変形されている。この状態において、前記表情分析402から表情分析データが送られてくると、表情パラメータ抽出403では、表情分析データを3次元モデル405a用の変形パラメータに変換し、その変形のための表情パラメータ404を最終的な符号化データとする。一方、復号化側では、まず、前記表情パラメータ404によって、3次元モデル405bを変形させることにより、3次元モデル405bの表情を変更する。即ち、図5におけるワイヤーフレームモデルの座標点に変換される。輝度・色付加407においては、変形されたワイヤーフレームモデルの各パッチ（線分によって囲まれる平面、あるいは曲面）に対応した、輝度、色等の変形前の画素情報をピックアップして、ワイヤフレームモデルにはりつけ、最終的な復号画像408を得る。さて、上記においては、符号化側から復号化側へ伝送される情報は3次元モデル変形用の表情パラメータ404のみでよく、極端に送信データの圧縮が実現できる。なお、これは、文字の場合にあてはめると、文字の原画像をそのまま送る場合と、文字を認識してそのコード情報を送る場合に相当する。

【0006】さらに、以上の従来例においては、画像のみについて説明したが、実際のシステム、例えば、テレビ電話、テレビ会議等においては、画像とともに、音声の伝送も不可欠であり、符号化方式は個々独立であっても、両者は切り放して考えるべきものではない。

【0007】

【発明が解決しようとする課題】従来の顔画像符号化方式は、以上のように処理されており、それだけでも、もちろん、有効な方式である。しかしながら、画像システムを構築する場合において、画像からは切り放して考えることのできない音声を、全く別のものとしてとらえ、両者、即ち、画像と音声の間にある相関性を全く利用していないため、さらなるデータ圧縮の高効率化を見逃してしまっている。相関の存在とは、即ち、さらなるデータ圧縮の可能性があることに等しい。

【0008】

【課題を解決するための手段】本発明は、音声の認識・合成符号化課程において発生する情報、即ち、音声認識情報を、画像の分析合成符号化課程における3次元モデ

ルの変形情報として利用する。

【0009】

【作用】2次元情報である画像からの表情解析処理の一部が、1次元情報である音声からの認識処理に置き代わるため、表情解析における処理時間が大幅に短縮される。

【0010】

【実施例】（実施例1）以下、この発明の実施例について図面を参照しながら説明する。

【0011】図1は本発明の第1の実施例における顔画像符号化方式の処理を示したブロック図である。図において、顔原画像401はカメラ、スキャナ等の画像入力装置から入力される人間の顔画像、表情分析402は、前記顔原画像401を入力データとし、表情を分析する。表情パラメータ変換403では、前記表情分析402の出力データ、および、後記口形状パラメータ105に基づき、3次元モデル405aを基準とした時の表情パラメータ404を抽出し、符号化データとして出力する。一方、復号化側では、まず、受け取った符号化データ、即ち、表情パラメータ404に基づき、モデル変形406において3次元モデル405bを変形させる。さらに、輝度・色付加407では、変形後のモデル画像に対して、画素毎に輝度・色をつけ、表示するための復号画像408が生成される。一方、音声系において、原音声101はマイク等から入力される人間の音声、特徴抽出102は前記原音声101から認識に必要な特徴量を抽出する。音素認識103では、前記特徴抽出102の出力をもとに、音素を認識する。音素認識結果106は口形状パラメータ変換104において、3次元モデルの口、あるいは、口周辺を変形させるための口形状パラメータ105に変換され、モデル変形406で使用される。一方、前記音素認識103における音素認識結果106は、音声合成107で逆に合成され、再生音声108として出力される。

【0012】次に動作について説明する。

【0013】図1において、画像系の処理順序は従来例の場合と殆ど同様であるので、個々の動作説明は省略する。以下、本発明が従来例と大きく異なる点、即ち、画像と音声の間に存在する相関性を利用して画像の符号化を行う点を中心に説明する。一般に、音声情報と画像情報は、それぞれ独立した情報源として考えられ、ゆえに、データ圧縮の分野においても、これまでは独立に符号化・復号化処理されていた。例えば、風景画面を音楽とともに映し出すテレビ画像を考えた場合、風景そのものと、同時に流される音声との間には、通常全く関連性はない。しかしながら、ある状況下においては、画像と音声の間に相関性が生ずる場合がある。例えば、人間が話している画像などが前記相関が生ずる場合に相当する。即ち、図2に示すように、無表情の時には閉じていた唇が、「あ」という音声を発する時には「あ」という口形

状に、また、「い」という音声を発する時には「い」という口形状に変形する。従って、音声情報から音素を認識すれば、その認識情報を使用して顔画像における口形状を変形させることができる。

【0014】図1において、特徴抽出102は、マイク等から入力された原音声101を、ある一定の間隔で周波数分析し、音素の特徴を抽出する。なお、ここで言う「音素」とは、音韻や音節などの総称である。また、特徴抽出102では、音声の有無を判別し、音声が存在する時のみ音素の特徴抽出を行う。次に、音素認識103では、前記特徴抽出102で得られた音素の特徴を、予め決めてある音素の標準パターンと比較して音素を認識する。音素の標準パターンは、個人差があるため、話者に応じて適応的に学習、書換えが行われる場合もある。音素認識結果106には、例えば、50音の他に継続時間、音の高低、強弱、あるいは、包絡線等の音韻的な情報も含まれる。こうして得られた音素認識結果106は、従来、表情分析402、および、表情パラメータ変換403の処理を経て抽出された表情パラメータ404の一部として、モデル変形406に送られ、3次元モデル405bを変形させる。輝度、色付加407において、変形後のモデルに輝度、色等の画素値をはり付け、復号画像408とするのは従来例の場合と全く同様である。また、音素認識結果106は、復号側に送られ、音声合成107において合成され、再生音声108となってスピーカ等から出力される。

【0015】(実施例2)図3は、本発明の第2の実施例における顔画像符号化方式の処理を示したブロック図である。

【0016】本実施例が上記第1の実施例と相違する点は、上記第1の実施例が、前記音素認識103、口形状パラメータ変換104の処理を経て抽出された口形状パラメータ105を、直接3次元モデルの変形に使用するのに対して、本実施例では口形状パラメータ105から、3次元モデルの口周辺を局部的に変形させ、原画像との差分をとり、この差分から3次元モデルの変形量、即ち、口周辺の表情パラメータ404の補正值を求めるようにした点である。図3において、口形状パラメータ105はモデル局部変形109によって、3次元モデルの口周辺を局部的に変形させ、顔原画像401との間で差分が取られる。この差分は実際の顔原画像401と、復号画像の間の差であるから、両者は小さい方が好ましい。そこで、既に求められている表情パラメータに対する補正值を、口形状パラメータ予測110において前記差分から予測し、補正後の表情パラメータを最終的な表情パラメータ404とする。以下、第1の実施例と同様な処理が行われる。

【0017】なお、以上説明した本発明による実施例の動作は、顔原画像に付随する音声情報が存在する場合のものであり、音声情報が存在しない場合には、顔の目や

鼻等の他の部分と同様、顔原画像から表情分析されて口形状パラメータが求められる。上記の動作の切り替えは、音声情報の有無を検出する特徴抽出102がこれを行う。

【0018】

【発明の効果】以上説明したように本発明によれば、2次元空間情報としての顔画像の表情分析に比べ、1次元情報として処理の容易な音声認識の結果を利用して、3次元モデルにおける口形状の変形を行うため、顔画像の符号化、具体的には、表情分析に要する時間が大幅に短縮される。さらに、第2の実施例では、上記の効果に加えて、より原画像に近い画像の再現が可能となる。

【0019】なお、上記では、3次元モデルを使用する場合を例にとり説明したが、2次元モデルであっても同様の効果が得られる。

【0020】また、上記本発明による実施例において、音声認識は、音声自体の符号化、伝送のために存在するものであるため、画像と音声の両方を扱うシステムであれば、負担になるものではない。あるいは、上記では、音声の符号化方式として音声認識、合成による方法を例にとったが、音声合成が存在せず音声認識処理のみの場合でも本発明の効果は変わらない。この場合、音声自体は別の符号化方式、たとえば、波形符号化等を用いて符号化してもよい。さらに、上記では、画像符号化方式として、分析合成符号化のみを使用する場合を例にとったが、従来の波形符号化方式と分析合成符号化方式を組み合わせた混合方式であっても、同様の効果が得られる。

【図面の簡単な説明】

【図1】 本発明の第1の実施例による顔画像符号化方式の処理ブロックを示した図である。

【図2】 本発明の第1の実施例による顔画像符号化方式において、音声と画像の相関を直感的に説明した図である。

【図3】 本発明の第2の実施例による顔画像符号化方式の処理ブロックを示した図である。

【図4】 従来例における顔画像符号化方式の処理ブロックを示した図である。

【図5】 従来例、および、本発明の実施例における顔画像符号化方式の3次元モデル用ワイヤーフレームモデルを示した図である。

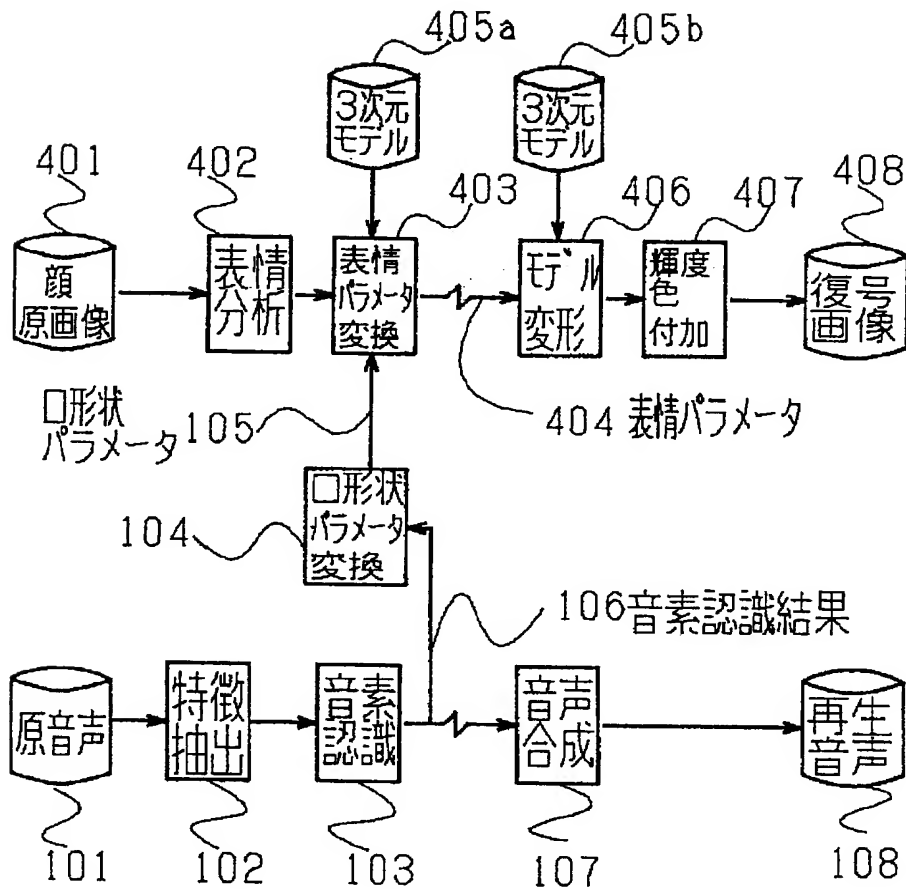
【符号の説明】

- 101 原音声
- 102 特徴抽出
- 103 音素認識
- 104 口形状パラメータ変換
- 105 口形状パラメータ
- 106 音素認識結果
- 107 音声合成
- 108 再生音声

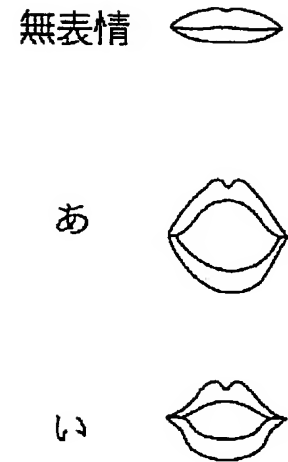
109 モデル局部変形
 110 口形状パラメータ予測
 401 顔原画像
 402 表情分析
 403 表情パラメータ変換

404 表情パラメータ
 405 3次元モデル
 406 モデル変形
 407 輝度・色付加
 408 復号画像

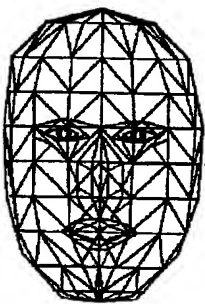
【図1】



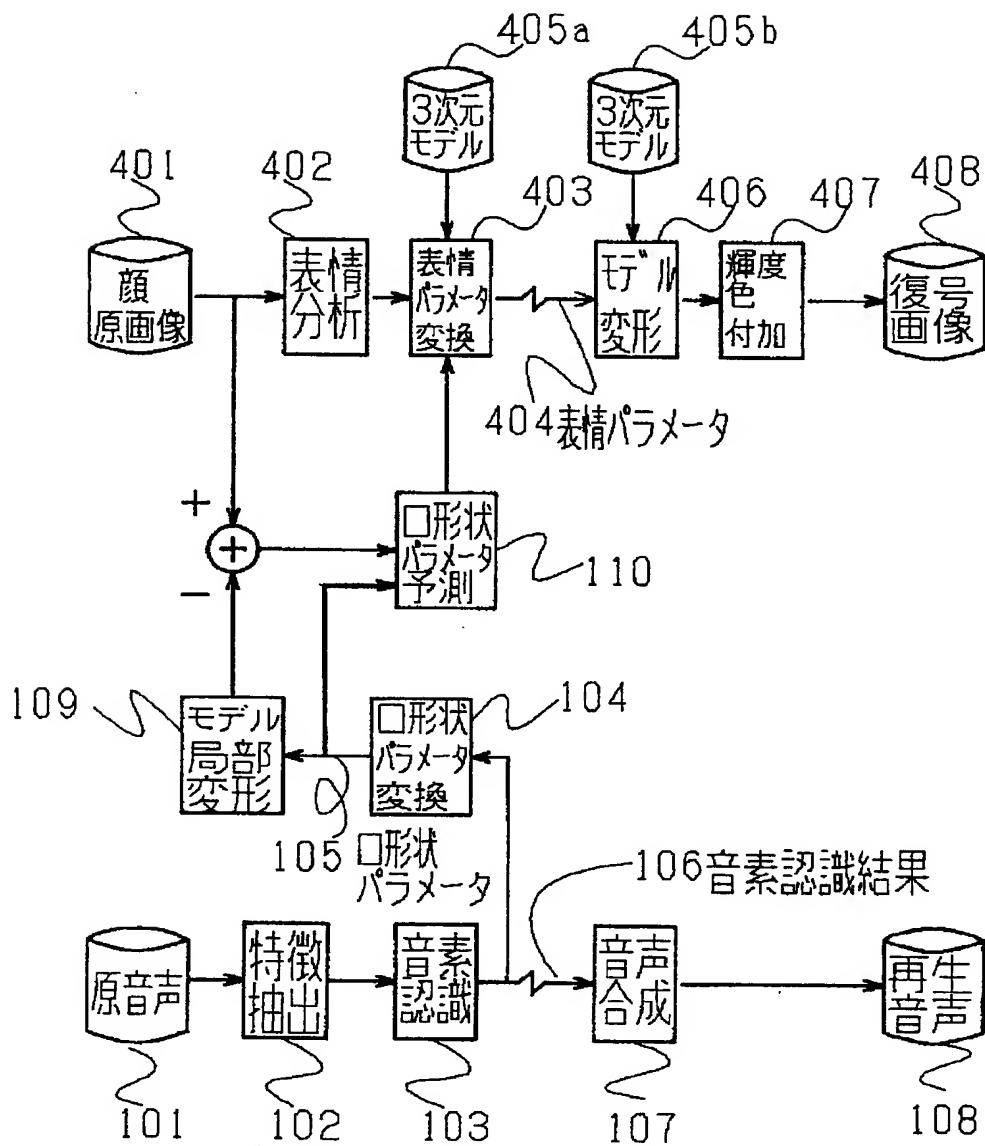
【図2】



【図5】



【図3】



【図4】

